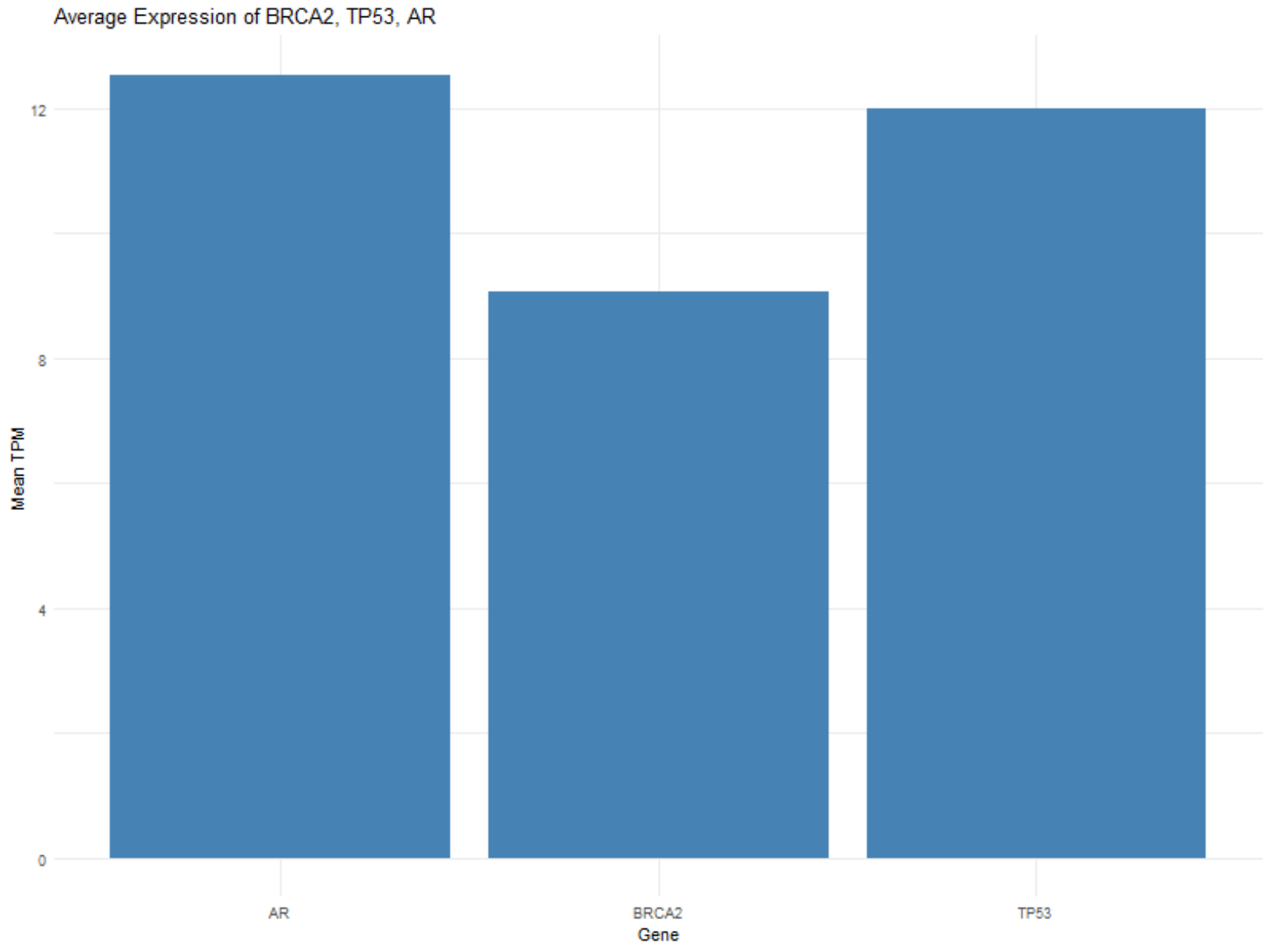
Clinical Action Plan Based on Gene Expression

BRCA2:
[.] Screen all breast cancer patients for BRCA2 mutations
[.] Prescribe Olaparib for eligible cases (HR+/HER2−)
[.] Refer to trial: KEYLYNK−009

TP53:
[.] Flag TP53 mutations in pathology reports
[.] Contact research coordinator for APR−246 trial

AR:
[.] Confirm androgen receptor status in prostate biopsies
[.] Enroll mCRPC patients in ARV−110 trial



---

title: "TCGA-BRCA Gene Expression Analysis Report" author: "Ayesha Gul" date: "2025-07-10" output: pdf_document: toc: true number_sections: true

---

1. Introduction

This report presents an analysis of TCGA-BRCA (Breast Invasive Carcinoma) RNA-Seq data, focusing on the expression of three clinically important genes: **BRCA2**, **TP53**, and **AR (Androgen Receptor)**. The data consists of 60,660 genes across 1,226 breast cancer samples obtained from TCGA (The Cancer Genome Atlas). The purpose of this study is to visualize gene expression and suggest clinical action points.

2. Methodology

2.1 Data Source

- File: `TCGA-BRCA.star_counts.tsv`
- Format: Gene-level STAR count expression matrix
- Rows: Ensembl gene IDs
- Columns: TCGA breast cancer patient sample IDs

2.2 Processing Steps

1. Removed version numbers from Ensembl IDs.
2. Filtered duplicate gene entries.
3. Annotated Ensembl IDs to gene symbols (using `org.Hs.eg.db`).
4. Extracted expression matrix for genes of interest: `BRCA2`, `TP53`, `AR`.
5. Calculated mean TPM (Transcripts Per Million).
6. Visualized expression via barplot and heatmap.

3. Results and Interpretation

3.1 Gene-Wise Summary

BRCA2
Expression varied widely across the cohort. Approximately half of the patients exhibited low BRCA2 expression (blue), while the remaining showed moderate to high levels (red). This variability may reflect underlying genomic heterogeneity or breast cancer subtypes.

TP53
TP53 showed consistent moderate to high expression (mostly red), aligning with its well-established role in tumor suppression. Nearly all patients showed similar expression profiles.

AR (Androgen Receptor)
AR expression was unexpectedly **high in most patients** (dark red), with a few low-expressing samples (blue). While AR is commonly studied in prostate cancer, its high expression here suggests a potential role in hormone-driven breast cancer or triple-negative breast cancer (TNBC) subtypes.

3.2 Clustering and Heatmap

A heatmap was generated to visualize gene expression patterns across all samples. Clustering revealed: - BRCA2: High inter-patient variability - TP53: Uniform expression - AR: Highly expressed in most patients with isolated low-expression clusters

Heatmap

4. Clinical Action Plan

BRCA2 Cases

- [X] Screen all breast cancer patients for BRCA2 mutations
- [X] Prescribe Olaparib for eligible cases (HR+/HER2-)
- [X] Refer to clinical trial: **KEYLYNK-009**

TP53 Cases

- [X] Flag TP53 mutations in pathology reports
- [X] Contact research coordinator for **APR-246** trial

AR Cases

- [X] Confirm androgen receptor status in biopsy reports
- [X] Enroll patients in **ARV-110** or similar trials if mCRPC or TNBC is observed

5. Skills Highlighted (CV/Portfolio Use)

- Bioinformatics pipeline development (RStudio)
- Data wrangling with `dplyr` and `readr`
- Gene annotation using Bioconductor packages
- Data visualization: `ggplot2`, `pheatmap`
- Report generation using `.Rmd` to `.PDF`
- Clinical interpretation of molecular data

6. Future Directions

- Perform DESeq2-based differential expression analysis.
- Stratify samples by PAM50 subtype or hormone receptor status.

- Integrate mutation data with expression.
- Explore pathways using GSEA or KEGG enrichment.

7. References

- TCGA Program (https://www.cancer.gov/tcga)
- Bioconductor packages: `org.Hs.eg.db`, `AnnotationDbi`, `pheatmap`
- Clinical trials database: https://clinicaltrials.gov/

This report was generated as part of a job portfolio project using real gene expression data and RStudio-based analysis.